CAN WEATHER FORECASTERS FORMULATE RELIABLE PROBABILITY FORECASTS OF PRECIPITATION AND TEMPERATURES?[1]

Allan H. Murphy
Advanced Study Program, National Center for Atmospheric Research[2], Boulder, Colorado

Robert L. Winkler
Graduate School of Business, Indiana University, Bloomington, Indiana

ABSTRACT

During the past ten years, many sets of subjective probability forecasts have been formulated on an operational or experimental basis by weather forecasters. In this paper we examine some forecasts prepared recently by National Weather Service forecasters. Samples of precipitation probability forecasts represent an operational program, while samples of probability forecasts of temperature represent an experimental program. An analysis of these forecasts indicates that weather forecasters can and do formulate reliable probability forecasts of precipitation and temperature.

1. INTRODUCTION

Since 1965, National Weather Service (NWS) forecasters have expressed their forecasts of precipitation occurrence in probabilistic terms, and these probability of precipitation (PoP) forecasts have been routinely appended to weather forecasts issued by the NWS. As a result of this nationwide program, several million PoP forecasts have been formulated and disseminated to the general public during the last ten years. These forecasts have been examined in some detail by NWS meteorologists (and others), and at least some of the results of these investigations have appeared in national and regional technical memoranda of the NWS (e.g., Cummings, 1971, 1974; Hughes, 1968, 1976; Roberts, Porter, and Cobb, 1969, Sadowski and Cobb, 1974). However, to the authors' knowledge no results related to the nationwide PoP forecasting program have appeared in the published literature in meteorology.

More recently, several experiments have been conducted in which NWS forecasters have expressed their uncertainty concerning forecasts of other variables in probabilistic terms. These variables have included maximum and minimum temperatures and the occurrence of precipitation in an area (PoP forecasts are point forecasts; see Section 3). While some of the results of these experiments have now appeared in various meteorological journals, many forecasters and operational meteorologists may not be familiar with these results.

The purpose of this paper is to briefly describe some recent results of the nationwide PoP program and some results of two experiments concerned with probabilistic temperature forecasting. While we are primarily concerned with the reliability of the forecasts, we will also consider their accuracy and skill. These attributes of the forecasts are defined and several measures of the attributes are identified in Section 2. In Section 3, we describe some recent results of the nationwide PoP forecasting program. Some results of two experiments in which NWS forecasters expressed the uncertainty in their forecasts of maximum and minimum temperatures in terms of credible intervals are summarized in Section 4. Section 5 consists of a summary and conclusion, including a brief discussion of the implications of these results for operational forecasting procedures and practices and for experimental programs.

2. DEFINITIONS AND SOME MEASURES OF THE RELIABILITY, ACCURACY, AND SKILL OF PROBABILITY FORECASTS

As indicated in Section 1, we are concerned in this paper with the reliability, accuracy and skill of probability forecasts. Reliability is defined as the degree of correspondence between forecast probabilities and observed relative frequencies over a set of forecasts. For example, if a forecaster used a probability of 30% for a PoP forecast on 50 different occasions, this set of 50 forecasts would be perfectly reliable if precipitation occurred on exactly 15 of the occasions. Any shift in either direction from a frequency of 15 (i.e., a relative frequency of 30%) represents a deviation from perfect reliability, although small shifts may occur by chance and should not be viewed as indications of serious difficulties. One reason that reliability is an important attribute of probability forecasts is because it is an indicator of the ability of forecasters, either collectively or individually, to quantify their uncertainty successfully.

Accuracy, on the other hand, is defined as

[1] *Supported in part by the National Science Foundation under Grants GA-41232 and ATM74-001610-A01.*
[2] *The National Center for Atmospheric Research is sponsored by the National Science Foundation.*

the average degree of correspondence between <u>individual</u> probability forecasts and the relevant observations. For example, a PoP forecast of 10% is viewed as more accurate than one of 30% if precipitation does not occur and less accurate if precipitation does occur. The Brier score (Brier, 1950) is generally used to "measure" this attribute of the forecasts. This score, or measure, is simply the mean square error of the forecasts.

The Brier score is also of interest because it can be partitioned into two terms (see Murphy, 1972), one of which is a measure of the reliability of probability forecasts. Thus, this partition provides a quantitative measure of the reliability of the forecasts. Specifically, the reliability term in the partition represents the average of the weighted squared differences between the forecast probabilities and corresponding observed relative frequencies, where the weights are the frequencies with which the respective probability values are used. The square root of this measure is the so-called "root mean square error in reliability," or simply the error in reliability.

Finally, skill is defined as the accuracy of the forecasts of concern relative to the accuracy of forecasts based solely upon the relevant climatological probabilities. This attribute is generally measured by a "skill score," which is simply the fractional improvement (in percent) of the Brier score for the forecasts formulated by the forecasters relative to the Brier score for the climatological forecasts. The higher the skill score, the more the forecaster has been able to improve upon climatology.

3. SUBJECTIVE PRECIPITATION PROBABILITY FACTORS

The subjective PoP forecasts formulated by NWS forecasters are point probability forecasts. Specifically, such a forecast expresses a forecaster's "degree of belief" that measurable precipitation (i.e., $\geq$ 0.01 inches) will occur during a specified period at a particular point in the forecast area (generally the official raingage). PoP forecasts are usually issued three or four times a day, and they consist of three probabilities which are valid for either three consecutive twelve-hour periods or a six-hour period followed by two twelve-hour periods. In this section we examine the reliability, accuracy, and skill of two different sets of PoP forecasts.

The first set of PoP forecasts was prepared by the forecasters at the NWS forecast office (WSFO) in Chicago, Illinois, during the period from July 1972 through June 1976. In formulating these forecasts, the forecasters were permitted to assign the probability values 0%, 2%, 5%, 10%, 20%, ..., 90%, and 100% to the occurrence of precipitation. A "reliability diagram," based upon all of the forecasts issued by the forecasters during this four-year period -- a total of 17,514 forecasts -- and the corresponding observations of precipitation, is presented in Fig. 1. The plotted points in the diagram represent the relative frequencies of precipitation when the various probability values were used. For example, the relative frequency of precipitation on those occasions on which the forecasters assessed a probability of 30% was 28.5%. We have also entered the number of forecasts next to each point on the diagram. Thus, forecasts of 30% were issued on 1574 occasions during the period, on 449 (28.5%) of which measurable precipitation actually occurred. The diagonal 45° line in this diagram represents perfect reliability, in the sense that the relative frequencies exactly equal the probabilities.

Examination of Fig. 1 reveals that these forecasts were very reliable. That is, the observed relative frequencies corresponded very closely to the forecast probabilities. However, a slight tendency did exist for the forecasters to overforecast (i.e., for the forecast probabilities to exceed the relative frequencies) for most probability values. In addition, it should be noted

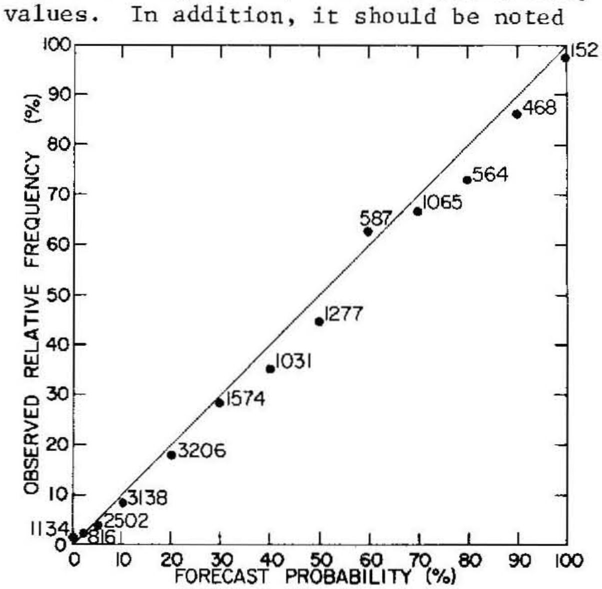


Figure 1. The reliability diagram for all of the PoP forecasts formulated by NWS forecasters at the Chicago WSFO during the period from July 1972 through June 1976.

that the frequency of use of the various probability values decreased, in general, as these values increased in magnitude. This latter result reflects the current state of the art in precipitation forecasting, in that it is simply not possible at present for forecasters to assign high probabilities, frequently and in a perfectly reliable manner, to an event which has an average climatological probability of occurrence of approximately 0.20. Nevertheless, the results presented in Fig. 1 indicate that the reliability of these subjective precipitation probability forecasts was excellent. The forecasters at the Chicago WSFO were clearly able to distinguish between those occasions on which, for example, probabilities of 25% and 30% should be assigned to the occurrence of precipitation.

In the previous paragraph we were concerned with the overall reliability of the forecasts. We might also ask, "How reliable were the forecasts formulated by individual forecasters?" To answer this question, we briefly examine the reliability of the forecasts prepared by the two forecasters who made the most forecasts during this period: namely, Forecaster A (2916 corecasts) and Forecaster B (2820 forecasts). The reliability diagram for Forecaster A is presented in Fig. 2. Forecaster A's forecasts were quite reliable, although he did exhibit a slight tendency to overforecast for most probability values. Note, however, that his 90% and 100% forecasts were remarkably reliable during this period. Fig. 3 represents the reliability diagram for Forecaster B. His forecasts were also quite reliable, and he did not exhibit as strong a tendency to overforecast as Forecaster A. These results indicate that the PoP forecasts formulated by individual forecasters were quite reliable.

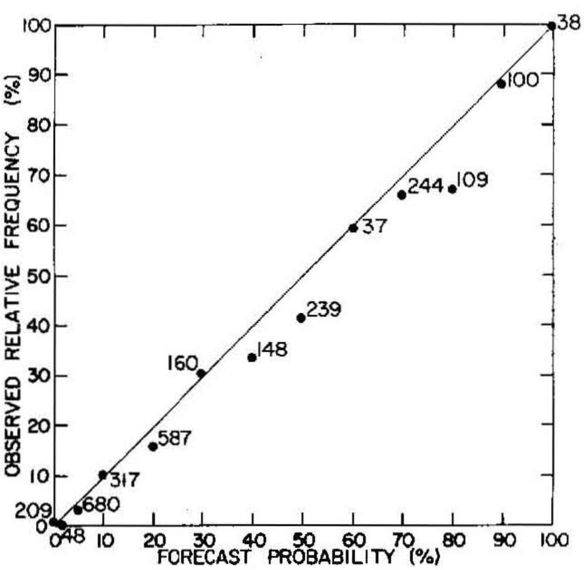


Figure 2. The reliability diagram for Forecaster A's PoP forecasts (a subset of the set of forecasts presented in Fig. 1.)

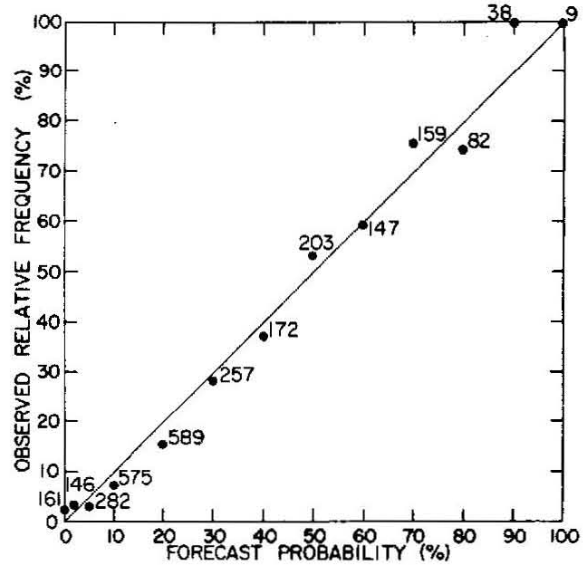
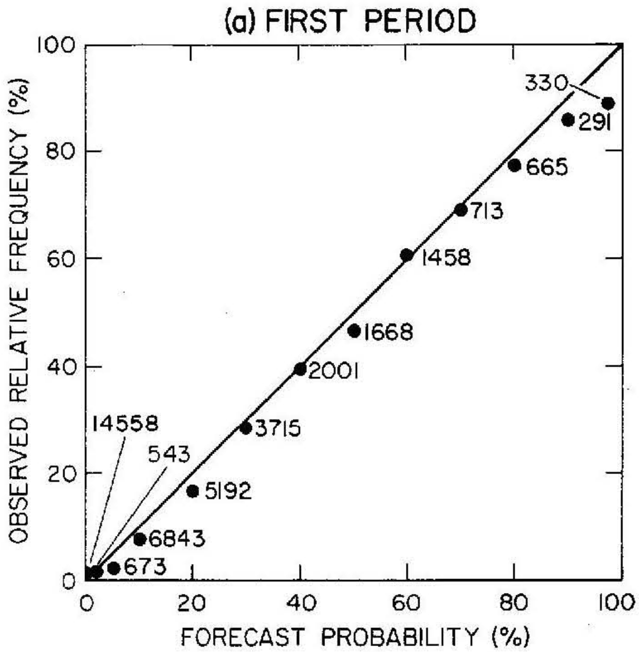


Figure 3. The reliability diagram for Forecaster B's PoP forecasts (a subset of the set of forecasts presented in Fig. 1.)

With regard to the reliability, accuracy, and skill of these forecasts, as determined by the measures described in Section 2, the Brier score for the entire set of forecasts was 0.130 (on a scale from zero to one, with zero being the best possible score). The Brier scores for Forecasters A and B were 0.124 and 0.128, respectively. The partition of the Brier score revealed that the error in reliability was 2.8% overall, and 4.4% and 3.5% for Forecasters A and B, respectively. Finally, the skill score for the entire set of forecasts was 30.9%; that is, the forecasts formulated by the forecasters represented a 30.9% improvement over forecasts based solely upon the climatological probabilities. Forecaster A's forecasts showed a 33.7% improvement, while Forecaster B's improvement was 32.3%. Thus, these quantitative results concerning the reliability of the forecasts, indicate that the PoP forecasts formulated by the NWS forecasters at the Chicago WSFO during this four-year period were reliable and skillful, both overall and for individual forecasters.

The second set of forecasts of concern consists of the twelve-hour PoP forecasts formulated by the forecasters in the Southern Region of the NWS during the period from April 1973 through March 1974 (see Cummings, 1974). The Southern Region comprises most of the eastern two-thirds of the southern half of the U.S. (ten states). Reliability diagrams for these forecasts are presented in Fig. 4 for (a) the first period (0-12 hours), (b) the second period (12-24 hours), and (c) the third period (24-36 hours). Each diagram is based upon approximately 38,650 forecasts. The highest probability value available to the forecasters in the Southern Region was > 95% (instead of 100%) and this value is plotted in the diagrams at 97.5%.



Examination of Fig. 4 indicates that these forecasts were quite reliable for all three periods or lead times. The first-period forecasts, in particular, were very reliable. A tendency did exist for the forecasters to overforecast slightly at the upper end of the probability scale, and this tendency increased as lead time increased. That is, the probability values were greater than the relative frequencies for the 80%, 90%, and > 95% forecasts in the first period and the tendency to overforecast appeared to extend down to the 70% and 50% forecasts in the second and third periods, respectively. In addition, it should be noted that, as in the case of the PoP forecasts formulated at the Chicago WSFO, the frequency of use of the probability values at the upper end of the scale was considerably less than that at the lower end of the scale, and this difference also increased as lead time increased. Once again, these results are simply an indication of the current state of the art in precipitation forecasting (in this regard, the average climatological point probability of precipitation in a twelve-hour period in the Southern Region is approximately 0.17). Nevertheless, the reliability of these subjective PoP forecasts was excellent for the first period, and it was still very good for the second and third periods except for those forecasts associated with the highest probability values.

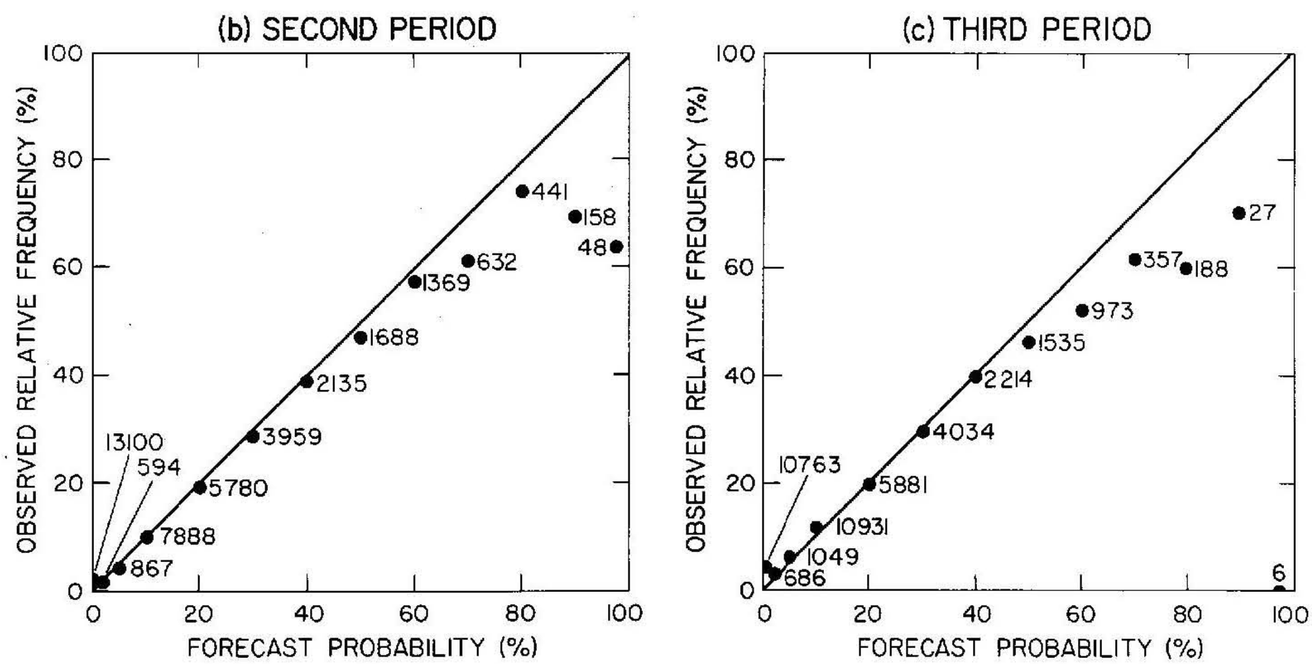


Figure 4. The Reliability diagram for the 12-hour PoP forecasts formulated by forecasters in the Southern Region of the NWS during the period April 1973 through March 1974. (a) First Period (0-12 hour) forecasts. (b) Second period (12-24 hour) forecasts. (c) Third period (24-36 hour) forecasts.

With regard to the reliability, accuracy, and skill of the Southern Region PoP forecasts, the errors in reliability -- as measured by the reliability term in the partition of the Brier score -- were 2.3%, 2.8%, and 4.4% for the first, second, and third periods, respectively. These errors are similar in magnitude to the errors in reliability in the Chicago PoP forecasts. The values of the skill score for the Southern Region forecasts revealed a 27.5%, 16.9%, and 8.3% improvement over climatology for the first, second and third periods, respectively. As expected, skill decreased as lead time increased. These results, when compared with the value of the skill score for the Chicago forecasts, indicate that the improvement over climatology for the Chicago forecasts was somewhat greater than that for the Southern Region forecasts (in this regard, the set of forecasts formulated at the Chicago WSFO included forecasts for all lead times). In any case, the PoP forecasts formulated by the Southern Region forecasters were also both reliable and skillful.

In summary, examination of these two sets of PoP forecasts, which were prepared in areas experiencing quite different weather regimes and precipitation types, provides convincing evidence that NWS forecasters can quantify the uncertainty inherent in their forecasts of the occurrence of measurable precipitation in a reliable and skillful manner.

4. CREDIBLE INTERNAL TEMPERATURE FORECASTS

In contrast to the forecasts of precipitation occurrence, forecasts of maximum and minimum temperatures are still expressed in categorical, or deterministic, terms. Specifically, temperature forecasts are generally either point forecasts (e.g., "the maximum temperature in Denver today will be 65°F") or interval forecasts (e.g., "the maximum temperature will be between 63°F and 67°F"). Point forecasts do not provide any information about the undertainty inherent in the forecasts, while interval forecasts provide only a very informal representation of the forecaster's uncertainty. With interval forecasts, the potential user of the forecast does not know whether the forecaster is almost certain that the maximum temperature will fall in the interval or whether there is only a 50-50 chance that the maximum temperature will fall in the interval

In an effort to determine whether forecasters could quantify the uncertainty in their maximum (high) and minimum (low) temperature forecasts, experiments were recently conducted at the WSFOs in Denver, Colorado, and Milwaukee, Wisconsin. The forecasters who participated in these experiments used cred-

Table 1. Relative frequency of occurrence of observed temperature below interval (BI), in interval (II), and above interval (AI) and average interval width for (a) variable-width forecasts and (b) climatological forecasts corresponding to variable-width forecasts.

Location	Set of Forecasts	Number of Forecasts	Percentage of observed temperatures: 50% intervals BI	50% II	50% AI	75% intervals BI	75% II	75% AI	Average width (standard deviation of width)(°F): 50% intervals	75% intervals
				(a)						
	All	132	25.8	45.5	28.8	10.6	73.5	15.9	6.2(1.3)	11.7(2.2)
Denver	Forecaster 1	64	29.7	37.5	32.8	9.4	76.6	14.1	5.8(1.3)	11.3(2.6)
	Forecaster 2	68	22.1	52.9	25.0	11.8	70.6	17.6	6.7(1.1)	12.0(1.7)
	All	432	18.1	53.9	28.0	8.1	79.4	12.5	5.9(2.0)	10.1(3.1)
Milwaukee	Forecaster 1	126	19.0	53.2	27.8	13.5	72.2	14.3	4.8(0.9)	8.1(1.3)
	Forecaster 2	171	12.9	59.1	28.1	4.1	82.5	13.5	6.5(2.5)	10.5(3.4)
	Forecaster 3	135	23.7	48.1	28.1	8.1	82.2	9.6	6.2(1.6)	11.3(2.9)
				(b)						
	All	132	31.1	44.7	24.2	18.9	65.2	15.9	14.8(4.2)	24.2(5.7)
Denver	Forecaster 1	64	32.8	39.1	28.1	21.9	59.4	18.8	15.5(4.1)	25.1(5.9)
	Forecaster 2	68	29.4	50.0	20.6	16.2	70.6	13.2	14.2(4.2)	23.2(5.4)
	All	432	17.8	56.9	25.2	8.1	81.7	10.2	14.5(3.9)	23.7(4.9)
Milwaukee	Forecaster 1	126	13.5	63.5	23.0	5.6	87.3	7.1	13.8(4.0)	22.5(5.0)
	Forecaster 2	171	18.7	57.9	23.4	8.2	79.5	12.3	14.7(4.1)	24.1(4.9)
	Forecaster 3	135	20.7	49.6	29.6	10.4	79.3	10.4	15.0(3.6)	24.2(4.6)

ible intervals to describe their uncertainty regarding forecasts of high and low temperatures. A credible interval temperature forecast is simply an interval forecast accompanied by a subjective probability which expresses the forecaster's degree of belief that the temperature of concern will fall in the interval (e.g., "the probability is 0.60 that the high temperature in Denver today will be between 63°F and 67°F"). Thus, credible intervals represent a straightforward extension of the interval forecasts frequently used in current temperature forecasting practice.

The Denver and Milwaukee experiments were conducted from August 1972 to March 1973 and from October 1974 to July 1975, respectively. Two forecasters in the Denver WSFO and three forecasters in the Milwaukee WSFO made 50% and 75% central credible interval forecasts of high and low temperatures during the respective periods. To obtain these intervals, the forecaster assessed the median, the 25th percentile, the 12-1/2th percentile, the 75th percentile, and the 87-1/2th percentile of his subjective probability distribution. Assessing each percentile involved an equal-odds indifference judgment -- that is, the division of an interval into two equally likely subintervals. For example, if 65°F is the forecaster's median temperature, then he believes that it is equally likely that the observed temperature will fall above or below 65°F. The 50% central credible interval is the interval from the 25th percentile to the 75th percentile, and the 75% central credible interval is the interval from the 12-1/2th percentile to the 87-1/2th percentile. All intervals in the experiments were assumed to include their end points, and all temperatures were recorded to the nearest degree.

Some of the results of these experiments are presented in Table 1a (for further results, see Murphy and Winkler, 1974, 1976). This table contains the relative frequencies, in percent, with which the observed temperatures fell below, in, and above the 50% and 75% credible intervals in the two experiments. Because these intervals are <u>central</u> credible intervals, the 50% intervals would be perfectly reliable if 25% of the observed temperatures fell below, 50% fell in, and 25% fell above the intervals. Similarly, the 75% intervals would be perfectly reliable if 12.5% of the temperatures fell below, 75% fell in, and 12.5% fell above the intervals. The results in Table 1a indicate that, overall, the intervals at both Denver and Milwaukee were quite reliable, in the sense that the forecast probabilities corresponded closely to the observed relative frequencies. A slight tendency did exist in both experiments for the relative frequency of observed temperatures above the intervals to be greater than that below the intervals.

Forecasts based upon climatological data can be used as a "standard of comparison" for the forecasters' credible intervals. The climatological forecasts considered here are variable-width intervals for high and low temperatures based upon historical data for the five-year periods immediately preceding the respective experiments, and they were computed on a monthly basis. These forecasts were analyzed in the same manner as the forecasters' intervals and the results are presented in Table 1b. An examination of the percentages of observations below, in, and above the intervals indicates that, overall, the climatological intervals were not as reliable as the forecasters' intervals. The tendency for the relative frequency of observed temperatures above the intervals to be greater than that below the intervals also existed for the climatological intervals at Milwaukee, but not for those at Denver. However, an examination of the climatological median temperatures in both experiments (not presented here; see Murphy and Winkler, 1976) reveals that these climatological medians tended to underestimate the observed temperatures at both Denver and Milwaukee. Thus, the forecasters' tendency to underestimate -- with respect to the variable-width intervals -- may be due in part to above normal temperatures during the experimental periods.

With regard to the individual forecasters, the tendency for the relative frequency of observed temperatures above the intervals to exceed that below the intervals was exhibited by all five forecasters, but this tendency was particularly pronounced only for Forecaster 2 at Milwaukee. Otherwise, the forecasters were generally quite close to the expected percentages of observed temperatures below, in, and above the intervals. One exception was Forecaster 1 at Denver, whose 50% intervals contained only 37.5% of the observations. However, an examination of the reliability of the corresponding climatological intervals reveals that only 39.1% of the observed temperatures fell in these intervals, indicating that "extreme" temperatures may have occurred more frequently than is normal on those occasions. Individually, the forecasters' intervals were, in general, slightly more reliable than the corresponding climatological intervals. Thus, the forecasters, on an individual as well as an overall basis, formulated quite reliable credible interval temperature forecasts.

Reliability is an important attribute of credible interval forecasts, and we have seen that the forecasters' intervals in the Denver and Milwaukee experiments were quite reliable. However, intervals based solely upon climatological data were only slightly less reliable than the forecasters' intervals in both experiments. Other attributes of these interval forecasts are also of interest, and precision is one such attribute. The average width of the intervals is a measure of their precision, and the widths of the forecasters' intervals and of the corresponding climatological intervals are also presented in Table 1. Overall, the average widths of the 50% and 75% intervals at Milwaukee were slightly narrower than the intervals at Denver. A comparison of the forecasters' intervals at both Denver and Milwaukee with the corresponding climatological intervals indicates that the former were less than half as wide as the latter. Individually, the forecasters' intervals, with Forecaster 1 at Milwaukee having particularly narrow intervals. Thus, the forecasters at both locations were able to formulate precise credible interval temperature forecasts.

In summary, the results of the Denver and Milwaukee experiments indicate that weather forecasters can quantify the uncertainty in their maximum and minimum temperature forecasts in terms of credible intervals in a reliable and precise manner, and that their forecasts were more reliable and much more precise than forecasts based upon climatological data.

5. SUMMARY AND CONCLUSION

In this paper we have analyzed subjective and probability forecasts prepared by weather forecasters under two considerably different sets of conditions. First, we examined samples of PoP forecasts, which have been formulated and disseminated to the public routinely by the NWS during the past ten years. Most weather forecasters have a considerable amount of experience at preparing PoP forecasts, and they also have access to objective guidance probabilities as well as other sources of information related to the likelihood of occurrence of precipitation. The two samples of PoP forecasts studied here were prepared in areas experiencing quite different weather regimes and precipitation types, and they indicate quite convincingly that PoP forecasts formulated by NWS forecasters are very reliable and skillful.

Next, we examined some probability forecasts of maximum and minimum temperatures. These forecasts were expressed in the form of intervals, accompanied by the forecaster's probability that the temperature of interest would fall in the interval in each case, and they were prepared on an experimental basis at two locations. The weather forecasters participating in the experiments had no previous experience at preparing such forecasts, and no similar guidance forecasts were available. Nevertheless, an analysis of the experimental forecasts indicate that they were very reliable and precise.

In summary, the results discussed here provide strong evidence that weather forecasters can do a very good job of expressing their uncertainty about precipitation occurrence and maximum and minimum temperatures in terms of probabilities. Furthermore, we feel that these results are indeed representative of the many sets of subjective probability forecasts that have been prepared on a routine or experimental basis by weather forecasters during the past ten years (e.g., see Murphy, 1976, for references to papers and reports describing other results). It appears, then, that we can provide an affirmative answer to the question posed in the title of this paper. Weather forecasters can and do formulate reliable probability forecasts of precipitation and temperature.

The results presented here have some implications both for operational forecasting procedures and practices and for experimental programs. With respect to operational programs, the PoP program appears to be quite successful and should be continued. Perhaps the most beneficial step that could be taken in terms of this program is in the area of educating the public. To make sure that the public understands PoP forecasts, we feel that the appropriate interpretation of a PoP forecast should be explained in a simple, clear fashion via the various media. Steps should also be taken to ensure that forecasters have the correct interpretation in mind when making PoP forecasts. The difficulties lie not just with the abstract notion of probability, but also with the specific probability used in a PoP forecast (i.e., a point probability) and with the definition of precipitation occurrence. Also, individuals who are skeptical about PoP forecasts might be more receptive to such forecasts if they knew that careful examination of large sets of forecasts (such as the results presented in this paper) indicate that they are very reliable and skillful. Thus, some publicity given to such results might be valuable.

In addition to the PoP program, consideration should be given to formulating other probability forecasts on an operational basis. The experiments concerning credible interval temperature forecasting suggest that probability forecasts of temperature might be formulated operationally, although careful thought needs to be given to the type of forecast that would be most appropriate. Instead of credible intervals, for instance, the probability that a temperature will be greater than or less than some particular threshold value might be of interest. Examples are the probability that the minimum temperature will be below freezing (or below some level that is critical for crops or plants) and the probability that the maximum temperature will be above some fixed level such as 90°F.

With respect to experimental programs, further experimentation concerning probabilistic weather forecasts seems warranted. Such experimentation could provide additional evidence regarding types of forecasts already studied experimentally (e.g., probability forecasts of temperature, point versus area precipitation probability forecasts) as well as information regarding types of forecasts not previously studied (e.g., probability forecasts of precipitation amount). It appears that weather forecasters are very good at expressing their uncertainty in terms of probability forecasts, and this ability should be utilized wherever possible to provide forecasts that could be helpful to the general public or to specific users.

6. ACKNOWLEDGEMENTS

We would like to acknowledge the cooperation of Allen D. Cummings and Lawrence A. Hughes, who provided the authors with information relating to the sets of PoP forecasts for the Southern Region of the NWS and for the Chicago WSFO, respectively.

7. REFERENCES

Brier, G.S., 1950: Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review*, V 78, pp 1-3.

Cummings A.D., 1971: Some Aspects of Probability of Precipitation Forecast Performance by WSFOs in the Southern Region - November 1969 through October 1970. *NOAA, National Weather Service, Technical Memorandum NWS SR-57*, 35 pp.

Cummings, A.D., 1974: Some Aspects of PoP Forecast Performance by Southern Region WSFOs - Revisted. *NOAA, National Weather Service, Technical Memorandum NWS Sr-77*, 18pp.

Hughes, L.A., 1968: Probability Verification Results (24 months), *ESSA, Weather Bureau, Technical Memorandum WBTM CR-19*, 29pp.

Hughes, L.A., 1976: On the Verification of Forecasts. *Preprints, Sixth Conference on Weather Forecasting and Analysis (Albany)*, American Meteorological Society, Boston, Mass., pp 1-8.

Murphy, A.H., 1972: Scalar and Vector Partitions of the Probability Score: Part I. Two-State Situation. *Journal of Applied Meteorology*, V 11, pp 273-282.

Murphy, A.H. 1976: Probability Forecasting in Meteorology: Status, Problems, and Prospects. *Preprints, Sixth Conference on Weather Forecasting and Analysis (Albany)*. American Meteorological Society, Boston, Mass., pp 9-12.

Murphy, A.H., and Winkler, R.L., 1976: The Use of Credible Intervals in Temperature Forecasting: some Experimental Results. *Decision Making and Change in Human Affairs* (H. Jungermann and G. de Zeeuw, Editors), Amsterdam, The Netherlands, D. Reidel, *in press*.

Roberts, C.F., Porter, J.M., and Cobb, G.F., 1969: Report on Weather Bureau Performance 1967-68 and Comparison with Previous Years. *ESSA, Weather Bureau, Technical Memorandum WBTM FCST-11*, 44pp.

Sadowski, A.F., and Cobb, G.F., 1974: National Weather Service April 1972 to March 1973 Public Forecast Verification Summary. *NOAA,*